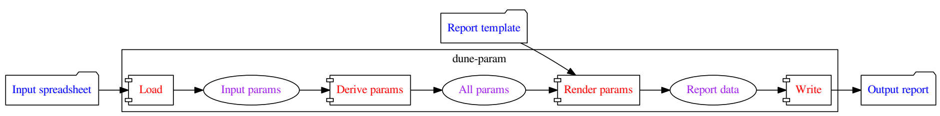
Annex 4B: Expected Data Rates for the DUNE Detectors

Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE)



August 14, 2015

DEEP UNDERGROUND NEUTRINO EXPERIMENT

1

Contents

1

List of Figures

1

List of Tables

1

Todo list

Chapter 1

Far Detector

1.1 Overview

A number of parameters from DUNE/LBNF requirements, design documents, and studies are used as input for estimations of the data rates in DUNE. In cases when derived parameters need to be generated based on these inputs this is accomplished using the software package *dune-params*[1] developed specifically for this purpose. As the inputs and estimations are refined the results presented in this annex can be regenerated at will.

Because the triggering and readout strategy and algorithms of the DAQ and analyses are under development and far from their final state of readiness, the data rate estimations involve some broad assumptions and leave open some choices. This is described in more detail in the following subsections.

1.2 Thresholds for the LArTPC Data Rate Estimations

There are three threshold levels considered for the purposes of characterizing the LArTPC data rates. These thresholds are assumed to be applied "per-wire" and on the basis of ADC values (which can be translated to units like MeV with proper calibration). Optimizing these thresholds for physics require additional study and are used here only to provide benchmarks for the data rate estimations. The data produced at each threshold are termed:

full-stream The full-stream (FS) threshold means there is no threshold at all. FS data is data where every time bin (as defined by the ADC clock) on every channel is read out.

zero-suppressed The zero-suppression (ZS) threshold is an ADC level roughly corresponding to what an 0.5 MeV deposition would produce in a single time bin on a single wire. All bins with ADC counts below this threshold are remove from the data stream.

high-energy A high-energy (HE) threshold is assumed which is such that all signals from radioactive decays are suppressed but low enough to not impact activity from beam neutrino interactions or potential nucleon decay activity. Further studies are needed to determine this threshold but currently it is taken, without rigor, to be 1 MeV–10 MeV.

1.3 Assumption on the DAQ

The data rate estimates make the following assumptions on the DAQ capabilities. In discussion with the DAQ experts they are expected to be satisfied. More information on the DAQ is available in the Volume 4 of the CDR.

- The DAQ trigger farm will be able to identify isolated activity collected in a single APA consistent with ^{39}Ar decay in order to suppress reading it to the DAQ data farm.
- The RCEs will be able to apply a zero suppression algorithms with a given threshold to data sent to the trigger farm.
- The RCEs will be able to apply trigger-specific zero suppression algorithms and thresholds to the data sent to the data farm.
- RCE-local data storage is left open as a possibility.

1.4 Sources of Data in the TPC

The data rate estimates are produced considering a specific sources of data.

in-spill Any activity in the detector which is coincident with the passage of beam neutrinos through the detector.

with-beam-ν A subset of *in-spill* where activity is consistent with a beam-ν interacting in the detector.

cosmic-μ Activity due to the passing of cosmic-ray muons through the detector.

radioactivity Activity due to the decay of radioactive isotopes.

atm-ν Activity consistent with interactions from atmospheric neutrinos.

noise Fluctuation in electronics noise (common-mode not considered).

1.5 Fundamental Parameters of the LArTPC

This section provides a selection of fundamental parameters used as input to the data rate estimations. The parameters are summarized in table 1.1

Table 1.1: The fundamental parameters serving as input to data rate estimations.

Parameter	Value
Full height	12.0 m
Full width	14.5 m
Full length	58.0 m
Detector modules	4
channel/APA	2560
APA/module	150
Active height (APA)	6.0 m
Active width (APA)	2.3 m
Drift distance	3.60 m
Drift velocity	1.6 mm μs^{-1}
Drift time	2.25 ms
bytes/sample	1.5 B
sample rate	2.0 MHz
# drifts/readout	2.4
readout time	5.4 ms
samples/readout	10,800

1.6 Full-stream Data

Full-stream (FS) data corresponds to reading all data in every ADC channel without application of any threshold. Estimating its rate is an exact calculation based on known parameters as it does not depend on the activity in the detector or the noise level of the electronics. As its name implies, it is the most voluminous type of data that can be generated by the TPC. The parameters which apply to this data are given in table 1.2.

The expected data rates for two scenarios of FS data are given in table 1.3. The first row gives the data size of one DAQ readout (5.4 ms). The second is appropriate for any strategy that intends to record FS data for each beam spill. The third contains two numbers that characterize data volume relevant to a strategy which aims to record FS data for Supernova Burst candidates. The final row shows the total annual data volume that the DUNE DAQ is capable of producing (in theory). These numbers are not meant to imply ongoing recording of full-stream data to permanent storage.

Table 1.2: Parameters pertaining to full-stream data rates.

Parameter	Value
Bytes per sample	1.5 B
DAQ sample rate	2.0 MHz
Channels per APA	2560
Number of APA per detector module	150
Number of modules	4
Total channels in DUNE	1,536,000
Drifts per readout	2.4
Drift time	2.25 ms
Beam spill repetition rate	0.83 Hz
Annual run time fraction	0.667

Table 1.3: Data volumes and rates for full-stream data acquisition.

Parameter	Value
Full-stream readout size	24.9 GB
Full-stream in-spill data rate	436 PB year^{-1}
Full-stream 1 second data volume	4.6 TB
Full-stream 1 minute data volume	276.5 TB
Full-stream 1 year data volume	145.4 EB

1.7 Zero-suppressed Data

There are options in choosing the exact zero-suppression (ZS) procedure, and the final choice has not been made. For these data rate estimates a very simple procedure is assumed: in each channel, all digitized time bins in which the ADC values are below the given threshold are removed. More discussion on possible alternative ZS methods and their impact on the data rate are give below.

The nominal ZS threshold is assumed to be the ADC counts produced when the equivalent of 0.5 MeV energy is deposited and the ionized wire is collected by a single wire. It is assumed that the application of zero-suppression at this threshold completely removes ADC counts due to just noise although an estimate of data rates due to noise is given.

Estimations of different sources of ZS data are summarized in table 1.5.

^{39}Ar Decays

The ^{39}Ar decay could potentially dominate the data volume. The end point of the ^{39}Ar decay is at 565 keV and about 25% of the beta spectrum is above the ZS threshold[**?**]. The expected total decay rate is 1.0 Hz kg^{-1}[**?**].

Table 1.4: Data rate estimations for ZS data from various sources. An additional FS data estimation is given for supernova burst (SNB).

Source	Event Rate	Event Size	Data Rate	Annual Data Volume
all ^{39}Ar	11.2 MHz	150 B	1.7 GB s^{-1}	53 PB
all in-spill				159 TB
with-beam-ν				79 GB
cosmic-μ	0.259 Hz	2.5 MB	647.4 kB s^{-1}	20 TB
beam-ν	8770 year^{-1}	2.5 MB	0.69 kB s^{-1}	22 GB
SNB cand. (ZS)	12 year^{-1}	16.7 GB	6366 B s^{-1}	201 GB
SNB cand. (FS)	12 year^{-1}	46.1 TB	17.5 MB s^{-1}	553 TB

s-volume

The ^{39}Ar ionization is very localized and it is assumed that any dispersion will not spread the charge out to any extent larger than a wire pitch nor the distance the charge will drift over one sample. Based on the expected shaping time that the electronics imposes on the signal the resulting waveform will be spread over 10 µs or 20 samples. Because offline signal processing is sensitive to tails of waveforms, even more time samples may be required. Finally, it is assumed a single collection wire and two induction wires in each view will register an appreciable signal.

As small as these events are, they are numerous enough that their data volume is not justified given their relative lack of physics importance. Some mitigation is required and will be developed.

The simplest mitigation is to increase the ZS threshold to be above the decay endpoint. This will produce a negative impact in removing small energy depositions associated with larger events and thus will only be considered if absolutely required to mitigate the rate.

A better approach is the one that drives the requirement on the DAQ that the trigger farm be capable of identifying isolated activity consistent with ^{39}Ar on a per-APA basis in order to veto its recording. The DAQ is expected to be able to provide this functionality. This then leaves ^{39}Ar which is accidentally coincident in the same APA with readouts from other activity such as beam-ν interaction and cosmic muons. The annual number of above ZS-threshold ^{39}Ar decay events coincident anywhere in the DUNE detector with beam-ν activity is given in table ref:zs-volume as 79 GB. Of that only 3% are coincident in the same APA bringing the added data rate to about 10% that of the beam-ν activity.

Supernova Burst

The Supernova Burst (SNB) data is estimated assuming a *false-positive* SNB rate of 12 year^{-1} and a readout time of 10.0 s. It should be emphasized that both these parameters are subject to modification and are used simply to provide benchmark examples.

On possible source of false positive SNB triggers may be an upward fluctuation in the rate of ^{39}Ar decays. Whatever initiates, it is assumed that the data volume of a false-positive SNB trigger is dominated by ^{39}Ar decays.

In the case of an actual SNB, its neutrino interactions will produce on order of 1000 events across the far detector modules over a time around ten seconds and with a neutrino spectrum up to 100 MeV. Given the importance of collecting SNB neutrinos, the same trigger-based reduction of ^{39}Ar will not be employed and thus it will dominate the data volume. These addition data volume due to the SNB events themselves is not significant.

Further, a lower ZS threshold may be considered for saving such candidate SNB occurrences. The possible data rate of SNB candidates is thus bound by the nominal ZS ^{39}Ar rate and the FS rate. The exact strategy for saving SNB candidates requires additional study and may have implications on DAQ hardware.

1.7.1 Noise

The nominal ZS threshold of 0.5 MeV is based on an older requirement of a 9:1 signal to noise ratio for a 1 MIP. The most likely MIP is 1.8 MeV/cm or 0.9 MeV per wire pitch. This implies an RMS noise requirement equivalent to 0.1 MeV and then the 0.5 MeV equivalent ZS threshold represents a 5σ cut. Across the 1,536,000 channels and the 10,800 per readout gives about 5000 samples above the nominal ZS threshold. Their distributed nature and isolated appearance make them subject to the same rejection criteria as isolated ^{39}Ar and thus can be ignored for the purposes of data volume estimates.

1.8 High-energy Threshold

For the purpose of these estimates the high-energy (HE) threshold is chosen at the level above the energy scale of the radiological backgrounds relevant for the LArTPC, and set at 10.0 MeV. A more careful study is needed to determine a potentially more optimal value for this threshold. The data rates with the HE threshold applied are summarized in table 1.6.

Table 1.5: Data rate estimations for data from activity above the high-energy (HE) threshold from various sources.

Source	Event Rate	Event Size	Data Rate	Annual Data Volume
^{39}Ar	-	-	-	-
cosmic-μ	0.259 Hz	2.5 MB	647.4 kB s^{-1}	20 TB
beam-ν	8770 year^{-1}	2.5 MB	0.69 kB s^{-1}	22 GB
SNB cand.	-	-	-	-

With the HE threshold in place, activity from the ^{39}Ar events and any SNB candidates will not be visible (i.e. will be rejected in DAQ). Although actual SNB events may result in neutrino with energies as high as 100 MeV applying such a high threshold to SNB candidates won't be optimal and is not being considered at this point. For this reason, contribution from SNB to these data is not calculated for this threshold setting.

1.9 Compression

The above estimates assume that the only compression employed is the lossy removal of data below some threshold. In addition, a lossless compression is possible. The DAQ can employ Huffman encoding during readout or the compression features provided by ROOT may be provided during file writing. In order to estimate the effects of compression, specific particle types and energies were simulated using LArSoft. The initial kinematics for the simulated samples covered a matrix of particles (e, μ, τ, π^+, π^-, π^0 and proton) and momenta (100 MeV to 6 GeV). The uncompressed and compressed sizes are summarized in figure **??**.

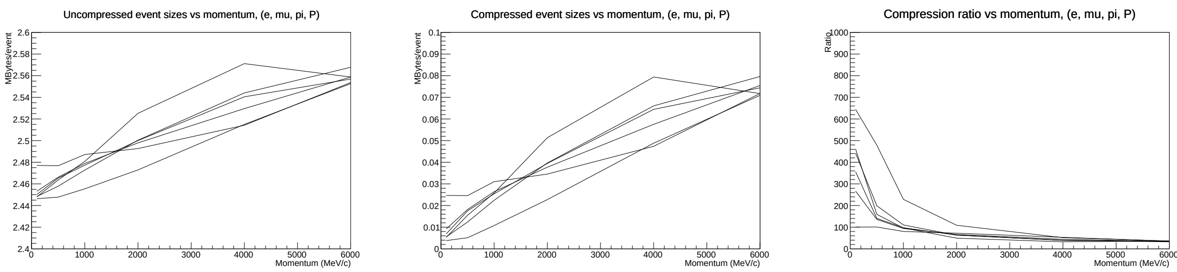


Figure 1.1: Estimation of DUNE zero-suppressed event sizes for different particles and momenta before (left) and after (center) compression and their ratios (right). Each line represents one particle type. The samples were generated with LArSoft assuming 10 kt detector module with 300,000 channels.

It must be noted that this exercise discovered the bulk of the compressed data written by LArSoft is actually made up channel ID numbers providing a data inflation of about a factor of 10. For uncompressed data this added about 50%. There are also additional data elements related in the "raw data" branch which are superfluous or at best repeated more frequently than is likely to be required. In the LArSoft output, even though zero suppression is applied, every channel has an associated a pedestal and a sigma as well as a compression number. There is also a sample count in addition to the one in the sample vector which is stored. These values do no vary yet each take up about the same (uncompressed) size as the ADC values and are significant in size even after compression. They are not included in figure **??**.

Finally, as can be seen in the figure, the change in uncompressed event size is only a minor function of event momenta and particle types. It is believed that LArSoft is saving additional overhead for all channels, even those that have been fully zero-suppressed.

The simulation used a detector module consisting of 300,000 wires and thus naively the event size in the full reference detector would be some 5 times larger. However, given the arguments above, it is expected that this leads to a gross over-estimate, possibly by one or two orders of magnitude from what should be expected if even just a little effort is put into designing a better storage schema. Once compression is assumed, the problems with this sub-optimal storage schema are lessened. We then select the uncompressed event size to be 2.5 MB and the compressed event size of 0.1 MB based on **??**. We identify this with being the zero-suppressed data produced by just activity due to beam interactions or similar energy events and accept the large over-estimation as a generous safety factor.

Chapter 2

Alternative Far Detector Design

2.1 Data rates tables for the alternate Far Detector Design

In the following the summary tables for the rates and data volumes, corresponding to the alternate Far Detector design, are reported. Given the amplification factor of the ionization electrons provided in the gas phase by the LEM detectors the signal to noise ratio can exceed 100:1 for a particle at the ionization minimum. Assuming conservative requirements on the argon purity, like 3 ms electrons lifetime, and a drift field of 0.5 kV/cm the minimal S/N after 12 m of drift path is in the range 12:1-60:1 depending on the LEM gain (20-100). Given the strips pitch of 3.125 mm, a noise level of 60:1 corresponds to about 11 keV. A zero-suppressed (ZS) threshold at $\approx 3\sigma$ above the mean noise level of the front-end electronics can be computed similarly to the one reported for the reference design. This corresponds to 33 keV threshold. The excellent S/N ratio provided by the double-phase amplification can translate in a very low energy threshold but also make easier the zero suppression since there are margins to increase the ZS threshold, still remaining at low energy, in the 100 KeV range. A higher threshold provides some immunity margins with respect to additional unforeseen noise sources which could be present on top on the intrinsic nose of the front-end amplifiers, like coherent noise related to grounding or external noise sources. This additional noise can quickly make less effective the ZS data volume reduction based on thresholds relying only on the expected front-end noise level. The DAQ system described in the alternate far detector design foresees, on top or independently than ZS, a lossless data compression scheme implemented in the digitization cards by using the Huffman compression algorithm. This compression factor is not taken into account in the full stream data volume calculation reported in the following.

Each 10 kton detector is read out by 80 CRP anode units of 3×3 m^2 with two perpendicular collection views segmented in strips of 3m length and 3.125 mm pitch. There are 1920 readout channels per CRP and 153600 channels per detector. Given the drift path of 12m the drift time amounts to 7.5 ms. The ADC sampling frequency is 2.5 MHz and its resolution is 12 bits. The number of ADC samples per drift window is then 18750.

These fundamental parameters serving as input for the data rate estimations are summarized in table 2.1

Table 2.1: The fundamental parameters serving as input to data rate estimations for the alternative far detector design.

Parameter	Value
Full height	12.0 m
Full width	12 m
Full length	60.0 m
Detectors	4
channel/CRP	1920
CRP/detector	80
Active height	12.0 m
Active width	12.0 m
Drift distance	12.00 m
Drift velocity	1.6 mm μs^{-1}
Drift time	7.5 ms
bytes/sample	1.5 B
sample rate	2.5 MHz
# drifts/readout	1.0
readout time	7.5 ms
samples/readout	18,750

The parameters which apply to the full-stream data rates are given in table 2.2.

Table 2.2: Parameters pertaining to full-stream data rates.

Parameter	Value
Bytes per sample	1.5 B
DAQ sample rate	2.5 MHz
Channels per CRP	1920
Number of CRP per detector	80
Number of detectors	4
Total channels in DUNE	614,400
Drifts per readout	1.0
Drift time	7.5 ms
Beam spill repetition rate	0.8 Hz
Annual run time fraction	0.667

The expected data rates for two scenarios of FS data are given in table 2.3.

Table 2.3: Data volumes and rates for full-stream data acquisition.

Parameter	Value
Full-stream readout size	16.1 GB
Full-stream in-spill data rate	268 PB year^{-1}
Full-stream 1 second data volume	2.1 TB
Full-stream 1 minute data volume	125.7 TB
Full-stream 1 year data volume	66.1 EB

n-volume

Chapter 3

Near Detector

3.1 Data Rate Estimations

This section summarizes the data rate estimations for the DUNE Near Detector (ND).

Table ?? give a summary of the event rates of different ND detectors and from different sources.

Table 3.1: Summary of DUNE Near Detector system masses, channel counts and rates.

Detector	Mass	Channels	ν-Event Rate
SST	8 t	215,040	1.6 Hz
ECAL	93 t	52,224	18.6 Hz
MuID	100 t	165,888	20.0 Hz

Table ?? give a summary of the event rates in the ND system from different sources.

Table 3.2: Summary of combined DUNE Near Detector system event rate estimations.

Source	Rate
Beam-ν	40.2 Hz
Rock	10.0 Hz
Cosmic	0.04 Hz
Total	50.2 Hz

All ND components will have a sample size of 5 B and given a nominal event size of 4000 samples the expected continuous data rate from the ND system as a whole is $1.0\,\mathrm{MB\,s^{-1}}$.

References

[1] https://github.com/DUNE/dune-params.